Week 11: Assignment 11

The due date for submitting this assignment has passed.

Due on 2025-04-09, 23:59 IST.

Assignment submitted on 2025-04-09, 22:16 IST

1) For which of the following problems are RNNs suitable? ***1 point***

- [ ] Generating a description from a given image.
- [x] Forecasting the weather for the next N days based on historical weather data.
- [x] Converting a speech waveform into text.
- [ ] Identifying all objects in a given image.

Partially Correct.
Score: 0.67

Accepted Answers:

Generating a description from a given image.

Forecasting the weather for the next N days based on historical weather data.

Converting a speech waveform into text.

2) Suppose that we need to develop an RNN model for sentiment classification. The input to the model is a sentence composed of five words and the output is the sentiments (positive or negative). Assume that each word is represented as a vector of length 100×1 and the output labels are one-hot encoded. Further, the state vector s_t is initialized with all zeros of size 30×1. How many parameters (including bias) are there in the network?

3992

Yes, the answer is correct.
Score: 1

Accepted Answers:

(Type: Numeric) 3992

3) Select the correct statements about GRUs **_1 point_**

- [x] GRUs have fewer parameters compared to LSTMs
- [x] GRUs use a single gate to control both input and forget mechanisms
- [ ] GRUs are less effective than LSTMs in handling long-term dependencies
- [ ] GRUs are a type of feedforward neural network

Yes, the answer is correct.
Score: 1

Accepted Answers:

GRUs have fewer parameters compared to LSTMs

GRUs use a single gate to control both input and forget mechanisms

4) What is the main advantage of using GRUs over traditional RNNs? **_1 point_**

- ○ They are simpler to implement
- ◉ They solve the vanishing gradient problem
- ○ They require less computational power
- ○ They can handle non-sequential data

Yes, the answer is correct.
Score: 1

Accepted Answers:

They solve the vanishing gradient problem

5) The statement that LSTM and GRU solves both the problem of vanishing and exploding gradients in RNN is **_1 point_**

- ○ True
- ◉ False

Yes, the answer is correct.
Score: 1

Accepted Answers:

False

6) What is the vanishing gradient problem in training RNNs? **1 point**

- The weights of the network converge to zero during training
- The gradients used for weight updates become too large
- The network becomes overfit to the training data
- The gradients used for weight updates become too small

Yes, the answer is correct.
Score: 1

Accepted Answers:
The gradients used for weight updates become too small

7) What is the role of the forget gate in an LSTM network? **1 point**

- To determine how much of the current input should be added to the cell state.
- To determine how much of the previous time step's cell state should be retained.
- To determine how much of the current cell state should be output.
- To determine how much of the current input should be output.

Yes, the answer is correct.
Score: 1

Accepted Answers:
To determine how much of the previous time step's cell state should be retained.

8) How does LSTM prevent the problem of vanishing gradients? **1 point**

- Different activation functions, such as ReLU, are used instead of sigmoid in LSTM.
- Gradients are normalized during backpropagation.
- The learning rate is increased in LSTM.
- Forget gates regulate the flow of gradients during backpropagation.

Yes, the answer is correct.
Score: 1

Accepted Answers:
Forget gates regulate the flow of gradients during backpropagation.

9) We are given an RNN with $||W|| = 2.5$. The activation function used in the RNN is logistic. What can we say about $\nabla = \left\| \frac{\partial s_{20}}{\partial s_1} \right\|$

1 point

- Value of ∇ is very high.
- Value of ∇ is close to 0
- Value of ∇ is 2.5
- Insufficient information to say anything.

Yes, the answer is correct.
Score: 1

Accepted Answers:
Value of ∇ is close to 0

10) Select the true statements about BPTT?

1 point

- [x] The gradients of Loss with respect to parameters are added across time steps
- [ ] The gradients of Loss with respect to parameters are subtracted across time steps
- [x] The gradient may vanish or explode, in general,if timesteps are too large
- [ ] The gradient may vanish or explode if timesteps are too small

Yes, the answer is correct.
Score: 1

Accepted Answers:
The gradients of Loss with respect to parameters are added across time steps
The gradient may vanish or explode, in general,if timesteps are too large